ORIGINAL PAPER

Predicting Airbnb pricing: a comparative analysis of artificial intelligence and traditional approaches

Nicola Camatti[1] · **Giacomo di Tollo**[2] · **Gianni Filograsso**[1] · **Sara Ghilardi**[2]

Abstract

Airbnb's distinctive model accommodates a broad spectrum of hosts, ranging from non-professionals to traditional establishments, resulting in a nuanced pricing system that poses challenges for prediction. This paper delves into Airbnb pricing and, to address this exercise, introduces computational approaches that combine traditional linear methods and advanced artificial intelligence techniques. Utilizing real data from the Netherlands, the study specifically focuses on all active Airbnb listings until September 2019. Our contribution stands out by incorporating the financial history of diverse rental offerings, a novel aspect compared to existing literature. Unlike other studies, our investigation spans various cities in the Netherlands, providing a comprehensive view beyond individual regions. Our results highlight the robust predictive capabilities of Artificial Intelligence techniques, that lead to lower susceptibility to overfitting and to superior overall performances. However, traditional methods also prove to be valuable, aiding in determining the significance of predictors in the predicting exercise. Collectively, these approaches offer valuable insights to enhance pricing strategies and overall performances within the Airbnb market.

✉ Nicola Camatti
  nicola.camatti@unive.it

  Giacomo di Tollo
  giditollo@unisannio.it

  Gianni Filograsso
  gianni.filograsso@unive.it

  Sara Ghilardi
  sara.ghilardi97@gmail.com

1 Department of Economics, Ca' Foscari University of Venice, San Giobbe Cannaregio 873, 30121 Venice, Italy

2 Department of Law, Economics, Management and Quantitative Methods (DEMM), University of Sannio, Via Delle Puglie 82, 82100 Benevento, Italy

Published online: 06 May 2024

⚛ Springer

1 Introduction

The hospitality platform Airbnb has been founded in 2008 and experienced an exponential growth ever since (Quattrone et al. 2018): 2.9 million people acted as hosts on Airbnb worldwide in 2022, and nearly one billion guests have been hosted since the community's creation, even though many scholars (and practitioners) expressed some criticism at the earlier phase of the platform development, due to some limitations intrinsic to peer-to-peer experiences (Grant 2013). The Airbnb platform is based on a two-sided market model, that serves as communication mean between accommodation-seeking tourists (guests) and accommodation-provider users (hosts) (Oskam 2019): these latter actors are partitioned into non-professional (i.e., actors for which renting the property represents a supplementary income and not a recurring business activity (Wegmann and Jiao 2017; Dolnicar 2019) and professional hosts (e.g., guesthouses, hostels, and bed and breakfasts (Oskam 2019; Mermet 2019)) that uses their real estate only for rental.

As for the real estates listed on the platform, different types of properties (listings) are shown on the community's website: single rooms, single houses, multiple rooms, and multiple houses (Adamiak 2022), partitioned into over 60 categories (as stated on Airbnb website, accessed on November 2023), defined to improve the guests' experiences, driven to use Airbnb due to some specific features (e.g., low costs, property qualities, novelty search, and authenticity (Gunter and Önder 2018; Guttentag et al. 2018)). As for the geographic factor, generally the distribution of accommodations in the online platform reflects the distribution of tourism demand and the traditional accommodation capacity (Adamiak 2022): Airbnb tends to be more prevalent in large cities (mostly experiencing single rooms) and leisure tourist destinations (mostly experiencing whole houses).

Due to its tremendous growth over the years, Airbnb has had disruptive effects in the hospitality industry, and has greatly affected the performance of real estate markets (Adamiak 2018). Furthermore, the Airbnb pricing processes may have a negative impact on the society, since the rented property would not be available for long-time renting, which is normally associated to the resident population, and would trigger an increase on the rental prices of a given city; on top of it, prices offered by Airbnb hosts are generally lower than the one offered by hotels operating in the same location, motivating owners to upload their properties on the platform and leading, again, to a negative impact on the rental market (Bernardi 2018; Guttentag 2015; Ball et al. 2014), and to increased probabilities of gentrification (Smith and Thorpe 2020) and segregation (Schelling 1969). In this framework, the need of determining the price of a listed real estate is a success key of the platform (Hung 2010), for both hosts (Rezazadeh et al. 2021) and guests (Zhang et al. 2017).

According to the platform developer (Hill 2015), the platform pricing exercise is performed by a machine learning based classifier conceived in 2012, aimed to classify whether a listing will get booked or not, and dynamically determines the daily prices based on changing market conditions and characteristics of listings.

Differently to other platforms such Uber and Lyft, the Airbnb pricing tool allows the users to decide whether they want to accept the proposed price, or to use their own pricing estimates: this leads to an increased complexity of a tentative pricing modelling exercise, that has to consider the market's characteristics and the specific hosts' management skills. Anyhow, it has been shown that the platform's pricing is not efficient, and leads to a revenue loss of roughly 46% (Gibbs 2018b), and this could be due to the challenges that sharing economy encounters when faced to the theory of profit maximization goals (McAfee and Te Velde 2006): traditional hospitality actors' pricing strategies depend on the need to generate profit based on accurate business information. Airbnb instead, pools together (also) hosts that are not professional actors, whose pricing could also be motivated by social reasons (Lampinen and Cheshire 2016), and that, unlike hotels, arbitrarily make their facilities unavailable in arbitrarily chosen dates, in which they do not want paying visitors, and in which any pricing exercise would make no sense (Dudás et al. 2017). In a nutshell, as reported by Gibbs (2018b), "Airbnb represents a new business model with non-professional management, instability regarding supply and potentially very inconsistent pricing", and starting from this consideration, our contribution aims to introduce an exercise of computational price modeling for the management of tourist rentals on Airbnb, based on both traditional (i.e. linear methods and generalized linear models) and artificial intelligence-based techniques, relying on predictors that are retrieved collecting and pre-processing available data on the Internet; we aim at assessing the significance of each predictor (i.e., by using linear regression and generalized linear models), along with using more black-box based methods (i.e., neural networks and random forests), whose pricing performances should be better, but that do not provide any tips about the predictors' significance, unless coupled with other techniques that are out of the scope of our contribution. Moreover, we combine the linear regression model with random forests and neural networks in an ensemble model, in order to gauge possible gains in terms of prediction accuracy. The predictors are combined in a voting ensemble, i.e. by averaging the individual predictions of models to form a final assessment. No accuracy gains are observed, suggesting that baseline linear predictors do not provide a valuable contribution, and that nonlinear models have a clear and unambiguous edge over linear ones.

In the selection of methodologies, we conducted preliminary experiments involving various models and data-driven approaches. Our emphasis was on interpretable methods, considering both model-driven and fully data-driven approaches. The objective was to compare these two types of methodologies, assess potential divergences, and evaluate the contribution of individual features.

Since the main drivers of the use of Airbnb are the listing price and its features (Sthapit and Jiménez-Barreto 2019; Guttentag et al. 2018), we want to understand whether there is a relation between these two components, and that's why we resort to traditional linear approaches: this will also provide evidence against the aforementioned assumption stating that Airbnb price is determined in a non-professional way, hence hindering prediction exercises. Furthermore, since the price can depend on spatial attributes, we have performed an analysis over different cities, in order to assess whether our findings can be generalized with regards to different geographical

locations. We point out that this additional test is aimed at evaluating different models for each municipality, avoiding comparisons across cities, since the sample sizes can differ significantly. Please notice that we are introducing some variables that represent the listing's (and host's) historic performances, and this is important because most literature does not consider the financial history of the listing, but, as it will be seen in what follows, they have a good significance in the introduced models. Please also notice that there is wide research on how spatial patterns impact the pricing exercise, especially with respect to the difference amongst neighborhoods and cities. So far, the outcome of this research is still contradictory (Sainaghi and Baggio 2020; Xu et al. 2019), and there is still lack of a significant amount of research over localities in Europe: that's why we have decided to apply our pricing methods, along with the whole country, to sets of data composed of listings belonging to single cities, in order to see whether any prediction differences arise, and also because pricing has a signaling effect towards the potential guests (Yao et al. 2019) and represent a proxy to assess the listing performances (Sainaghi and Baggio 2020). For all the aforementioned reasons, a pricing exercise is important for both tentative host and guest, along with policy makers, civil society, related industry, and academics.

Our paper is organized as follows: Sect. 2 discusses about the determinants of the Airbnb prices; Sect. 3 outlines the related literature about Airbnb listings' pricing. Section 4 introduces our computational approach and Sect. 5 discusses its results, before concluding and proposing further extensions in Sect. 6.

2 The determinants of Airbnb pricing

Numerous research contributions have attempted to unravel the complexities of Airbnb pricing, each employing distinct variables as predictors. In this section, we delineate the variables proposed in the relevant literature, accompanied by suggested guidelines for their partition.

Understanding the determinants of prices on Airbnb is pivotal, given specific characteristics of the platform, including: the small size of accommodation providers and their (often) not professionally trained pricing behaviors (Stabrowski 2017; Chen and Xie 2017); the flexibility of hosts to choose the rental price and period, without having to comply with service offering obligations (Sainaghi 2020); the influence of reviews left in the online platform, which is higher than the one experienced in the hospitality industry (Fu et al. 2021). Research on the topic reveal that Airbnb hosts do not apply dynamic (or aggressive) pricing strategies (Gibbs et al. 2018b),and that their risk-assessment procedures are highly influenced by communication, reputation, and trust (Sthapit and Jiménez-Barreto 2019). In this direction, a special attention has been paid to understanding what features have an impact on the price related to the specific listing, with special emphasis on trust-related aspects (e.g., the customers' references): several studies have reported that good reputation leads to an increase of product sales (Chevalier and Mayzlin 2006), but this could not be the case of Airbnb, in which the sellers (hosts) can sell only a given quantity of scarce resources (i.e., an apartment can at most be rented out 365 nights per year):

Table 1 Variables used for Airbnb pricing in the related literature.

Variable name	Category	Literature
Host attributes	Platform-specific	Wang and Nicolau (2017); Quattrone et al. (2018); Tong and Gunter (2020); Kwok & Xie (2019); Xie et al. (2019a, b)
Online review ratings	Platform-specific	Wang and Nicolau (2017); Chica-Olmo et al. (2020); Tong and Gunter (2020); Gutt and Herrmann (2015); Ikkala and Lampinen, (2014); Moreno-Izquierdo et al. (2019)
Number of listings in the area	Platform-specific	Chica-Olmo et al. (2020); Wang and Nicolau (2017); Gibbs et al. (2018a, b); Cai et al. (2019b)
Cleaning fee	Platform-specific	Gunter and Önder (2018); Perez-Sanchez et al. (2018); Lorde (2019)
Property location	Location-based	Wang and Nicolau (2017); Cai et al. (2019a); Chattopadhyay and Mitra (2019); Kwok and Xie (2019); Lorde (2019); Moreno-Izquierdo et al. (2019); Tong and Gunter (2020); Chen and Xie (2017); Hossain (2020); Gibbs et al. (2018a, b); Perez-Sanchez et al. (2018)
Distance to city center	Location-based	Chica-Olmo et al. (2020); Xu et al. (2019); Quattrone et al. (2018); Wegmann and Jiao (2017); Wang and Nicolau (2017)
Bathrooms	Location-based	Tong and Gunter (2020); Chica-Olmo et al. (2020); Xie et al. (2019a, b); Moreno-Izquierdo et al. (2019); Kwok and Xie (2019)
Bedrooms	Location-based	Ert et al. (2016); Chen and Xie (2017); Tong and Gunter (2020); Chica-Olmo et al. (2020); Xie et al. (2019a, b)
Availability of parking	Location-based	Wang and Nicolau (2017)
Site attributes	Location-based	Wang and Nicolau (2017)
Property type	Location-based	Xu et al. (2019); Wang and Nicolau (2017); Gutt and Herrmann (2015)
Rental rules	Location-based	Wang and Nicolau (2017)
Number of bus stops	Location-based	Xu et al. (2019); Quattrone et al. (2018); Türk et al. (2021)
Distance to trains	Location-based	Türk et al. (2021); Adamiak et al. (2019); Falk et al. (2019)
Distance to airport	Location-based	Türk et al. (2021); Adamiak et al. (2019); Falk et al. (2019)
Number of hotels	Location-based	Wang and Nicolau (2017); Quattrone et al. (2018)
Points of interest	Location-based	Xu et al. (2019); Tong and Gunter (2020); Quattrone et al. (2018)
Sightseeing	Location-based	Xu et al. (2019); Quattrone et al. (2018)
Number of restaurants nearby	Location-based	Xu et al. (2019); Perez-Sanchez et al. (2018); Quattrone et al. (2018)

Table 1 (continued)

Variable name	Category	Literature
Shopping centers nearby	Location-based	Xu et al. (2019); Perez-Sanchez et al. (2018); Quattrone et al. (2018)
Nightlife	Location-based	Xu et al. (2019); Perez-Sanchez et al. (2018)
Supermarkets	Location-based	Önder et al. (2018); Quattrone et al. (2018); Perez-Sanchez et al. (2018)
Listing density	Social	Cai et al. (2019a); Debooser et al. (2019); Kwok and Xie (2019); Tang et al. (2019)
Population density	Social	Quattrone et al. (2018); Debooser et al. (2019); Moreno-Izquierdo et al. (2019); Tang et al. (2019); Eugenio-Martin et al. (2019)
University	Social	Xu et al. (2019)
Green space	Location-based	Falk et al. (2019); Xu et al. (2019); Adamiak et al. (2019)
Water	Economic	Falk et al. (2019); Xu et al. (2019); Adamiak et al. (2019); Türk et al. (2021)

Category can be instantiated to platform – specific, location – based, social, and economic, as discussed by Quattrone et al. (2018)

in this case, an increase in requests due to high reputation can only lead to higher prices (Gutt and Herrmann 2015).

Several studies delve into the factors shaping Airbnb prices (see Table 1 for an overview of the main variables used for Airbnb pricing in related literature), and their conclusions directly align with research on the broader hospitality industry (Chen and Xie 2017), integrating considerations of the Airbnb digital features (Wang and Nicolau 2017). These analyses assess a comprehensive array of features, with a consensus emerging on the most impactful determinants. Studies, such as (Gibbs and al. 2018a), outline that room properties, listing position, and host specifics wield significant influence over listing prices. Additionally, Li et al (2016) identifies other critical factors, including distance from points of interest and city landmarks. The study by Zhang et al. (2017) introduces platform-based features, encompassing the number of reviews, rating, and listing age, while incorporating two geographical variables: distance to the city center and the nearest highway. Notably, it reveals a negative correlation between distance to the city center and price and explores the multifaceted effects of proximity to the nearest highway. Lawani et al. (2019) emphasizes the limited impact of neighborhood features, such as distance to the city center, train station, or violent crime areas, underscoring the importance of reviews and room qualities. Location of listings was also considered by Xu et al. (2019) as a determinant of Airbnb prices. Certain characteristics of the neighborhood where the accommodation is located (such as water, presence of vegetation, art and human landscape, travel and transportation, universities, and nightlife) emerged as important factors influencing Airbnb prices. Finally, regarding the location of listings, according to Zhang et al. (2017) Airbnb prices are influenced by the proximity of accommodations to certain points of interest. Other works investigate the effect of seasonality (Deboosere et al. 2019; Quattrone et al. 2018), in relation to the location of listings, including the characteristics of the neighborhood and the available services.

Different groups of variables are used to describe the different features of a property, such as physical characteristic, accessibility, socioeconomic attributes, and environment-related issues (Johnson 2003; Čeh et al. 2018; Anselin 1988). According to (Sainaghi 2021), six blocks of variables can be identified in order of intensity of influence on Airbnb's prices: listing variables, guest characteristics, location, guest reviews, destination characteristics, and external comparison. Alternative approaches to categorizing Airbnb price determinants were investigated by (Wang and Nicolau 2017). Within these classifications, the factors that exert the most significant influence on prices include host attributes, site and property attributes, amenities and services, rental rules, and online review ratings.

According to Perez-Sanchez et al. (2018) daily listing prices are influenced by several factors, including accommodation attributes, listing and host characteristics, tourism-related environmental characteristics, and listing location. All these factors can be summarized in three spheres of concern: physical characteristics, factors influencing users' perceptions, and location. Quattrone et al. (2018) reviewed the use and importance of both geographic and non-geographic variables used in the specific case of Airbnb price prediction, partitioning the observed variables into platform-specific variables (i.e., number of Airbnb listing in a given area),

location-related variables (i.e., distance to center, points of interest and public transports nearby, presence of hotel and population density), social variables (i.e., race diversity index, income diversity index, proportion of people employed in the art sector, talent index and presence of young people), and economic variables (unemployment rate, poverty and income metrics, median value of an household in the area, and proportions of dwellings that are occupied by its owners).

In order to study the determinants that influence the price of Airbnb accommodations, Perez-Sanchez et al. (2018) used a multivariable analysis technique. The variables were then analyzed by adopting both ordinary least squares and quantile regression methods. In addition, Teubner et al. (2017) has investigated the determinants of Airbnb listing prices, considering factors such as average rating score, number of ratings, Superhost status, identity verification, photos, and membership length. Through analysis employing a hedonic price regression model, Teubner et al. (2017) asserts that the host's rating score, the duration of the host's membership on the platform, and the number of photographs of the accommodation are the most influential variables affecting Airbnb accommodation prices.

It is important to note that the literature has considered other variables as well, including rental capacity, breakfast availability, pool access, Internet/television, washer/dryer facilities, kitchen amenities, and distance from the coast. However, providing a comprehensive overview of all variables used in the related literature goes beyond the scope of our contribution. Interested readers are directed to studies by Quattrone et al. (2018), Xu et al. (2019) and Wang and Nicolau (2017). It's also essential to acknowledge that while the concept of the sharing economy is not new (Albors et al. 2008; Sundararajan 2017), the introduction of large-scale mediating technologies has propelled the commercial success of internet-based platforms (Zervas et al. 2017). This has garnered attention from academics and policymakers, often relying on the assumption that their focus on sharing technologies differentiates them from traditional markets (Hamari et al. 2016). Airbnb is frequently cited as an exemplar of this new paradigm (Mikhalkina and Cabantous 2015; Sutherland and Jarrahi 2018), leading to numerous contributions exploring the emergence of new online discriminants capturing customer experiences, such as the presence of online pictures (Fagerstrøm et al. 2017) or behaviors stemming from racial discrimination (Schor and Attwood-Charles 2017).

Importantly, most literature does not delve into the financial history of the listings, despite being a crucial aspect in reflecting key issues such as hosts' attitudes towards managing their rental offers and achieving their results of revenue (Karlsson & Dolnicar 2016; Visser et al. 2017). In our contribution, however, we introduce the revenues from the last twelve months (referred to as LTM), along with other attributes not yet widely available in the literature. These attributes also convey the hosts' commitment to constantly improving their real estate business (Rubino & Coscia 2018), such as the number of bookings (Van der Borg et al. 2017), the days available on the agenda, and the number of days blocked (Deboosere et al. 2019; Gunter & Önder 2018; AirDNA LLC, 2018). Various partitioning approaches have been proposed in the literature, such as (Perez-Sanchez et al. 2018), introducing accommodation characteristics, host features, and environmental characteristics, or (Teubner et al. 2017), which categorizes variables into internal and external (see Sainaghi

(2021) for more detais). Furthermore, different approaches have focused on defining the relationships between prices and aspects related to the specific characteristics of the economy of a given territory, especially in determining the effect of Airbnb on specific regions or municipalities. In this study, we experiment with price prediction exercises that include a plurality of cities as well as the entire country (i.e., the Netherlands).

3 Methods for Airbnb pricing

A common research goal in the related literature is to investigate the direct (or indirect) contribution of the variables reported in the previous section on Airbnb prices. To this extent, it has been shown that variables associated to the listings' size are the most apt to explain the price variance (Cai et al. 2019b; Ert et al. 2016).

Early Airbnb pricing models draw inspiration from existing approaches in real-estate pricing literature. linear models (Li et al. 2016), for instance, have enjoyed widespread use, with many contributions employing ordinary least squares (OLS) linear regression (Čeh et al. 2018). These models are often compared to Multi-Scale Affinity Propagation (Zhang et al. 2017; Lawani et al. 2019) and Hedonic pricing models, that operate under the assumption that an object's price is constructed by its features, allowing the regression of a listing's price on its features to evaluate the contribution of each feature to the overall price (Čeh et al. 2018). Major empirical issues, inherent limitations, and advantages of these models are discussed by (Chau and Chin 2003), shedding light on their applications to the housing market: they provide an annotated bibliography delving into the effects of locational, structural, and neighborhood-related housing attributes. Intriguingly, they identify the attributes used in the estimation of hedonic price models and their varying effects on housing prices, contingent on the market.

Many Airbnb pricing studies rely on extensive datasets of listings and hedonic (linear) regression models (Guttentag 2019; Tang et al. 2019), and they show a drawback in the assumption of linearity, which is not guaranteed to hold in the Airbnb scenario (Chattopadhyay and Mitra 2019; Chumney and Simpson 2006); furthermore, multicollinearity poses challenges for large-scale application (Limsombunchao 2004). These considerations have prompted a shift toward more sophisticated methods that eschew model assumptions.

Machine learning (Rezazadeh et al. 2021; Park and Bae 2015), text mining (Tang and Sangani 2015), fuzzy methods (Andria et al. 2021), and other approaches aimed at addressing the Revenue Management problem (i.e., predicting pricing to maximize profits in a scenario of scarce resources (Dana 2008)) have gained attraction, as well as similar investigations in the hospitality market concerning hotel booking prices (Harewood 2006).

Neural networks have been utilized by Rezazadeh et al. (2021) in comparison to linear regression and other intelligent techniques. Additionally, Luo et al. (2019) employed neural networks, comparing them to a supervised learning algorithm that implements the gradient boosted trees algorithm (XGBoost, also utilized by Cai et al. 2019a, Kokasih and Paramita 2020), Random Forest, as well as linear

and unweighted K-nearest neighbor. Comparative works have also explored linear/hedonic models in contrast to artificial intelligence-based approaches, as seen in the work of Park and Bae (2015).

4 Methodology

In this section, we briefly review the basics of linear and non-linear models for prediction. In this contribution, the main idea is to characterize the problem of Airbnb pricing in a novel way, based on well-known machine learning algorithms, which implement purely data-driven approaches, and compare this with models that assume models underlying data. In this latter case, we want to evaluate how different model assumptions would affect the predictive performance, e.g., non-linearity in the relationship might imply a substantially lower performance of the baseline linear regression relative to other models. In this framework, heteroskedasticity can be managed through Generalized Linear Models (GLMs), whereas Decision Tree can provide a minimal benchmark for non-linear models. Finally, random forests constitute our baseline ensemble model: compared to Decision Tree, this algorithm should better handle large datasets and provide robustness to noise. Similarly, we also want to understand whether a complex and highly parameterized model, such as the artificial neural network, can add value to our forecasting exercise. Further extensions to more recent algorithms go beyond the scope of this paper and are left for future work. The section is organized as follows. In Sect. 4.1 we focus on the family of linear models, then in Sect. 4.2 we describe decision trees and random forests; finally, in Sect. 4.3 we discuss artificial neural networks, along with a brief description of the procedure used for training.

4.1 Linear regression and generalized linear models

Linear regression models are widely used in many research areas, due to their ease of use and generality: their aim is to explain a response variable by means of a linear relationship amongst a set of predictors, also considering a normally distributed error term, and their application to Airbnb pricing have been discussed in Sect. 3. When using linear regression, there is no limit to the value attained by the response variable, and this could lead to drawbacks when the range of the response is restricted: in these cases the relationship between the response variable and the predictors is not necessarily a simple linear form, hence one has to resort to a different (and more general) model. GLM rely on a linear relationship amongst the predictors, but they assume that the distribution of the response variable is non-normal; furthermore, the relation between the mean of the response variable and the linear combination of predictors is defined by using a suitable link function (usually employing a logit, probit or inverse function). In this way, it is possible to model the response using one of the various probability distributions pertaining to the exponential family GLMs represent nowadays a well-known tool for statistical analysis: we have tested several GLMs in order to determine which one is the most suitable

for the problem at hand, and in what follows we report the outcomes obtained by the Gamma model with the Gamma variance function and the inverse link function, which is well-suited when dealing with strictly positive continuously distributed data (which is our case over many variables) and are easier to interpret.

4.2 Decision trees and random forests

Decision trees are machine learning algorithms that use predefined criteria to recursively partition the set of data into two groups, until a stop condition is met. They have been used for both prediction and classification tasks, and both applications are based on splitting an internal tree node, relying on a subset of predictor variables which is selected depending on predetermined splitting criteria (e.g., entropy).

Decision trees show the drawback of being sensitive to overfitting, hence some guidelines have been introduced to overcome this phenomenon: one of those consist on (recursively) considering only a subset of the observations and build many individual trees: this is the base mechanism of Random Forests, which are machine-learning algorithm for prediction based on a combination of tree predictors, in which each tree is built based on the values of a random vector, which is generated by using the same distribution for all trees: they have shown good generalization errors, that depends on the individual trees fitness, but also on their correlation.

Random forests overcome the limitations of decision trees: they consist of several decisions trees and use binary splits to predict the model's output: they use random training datasets, and random predictors to predict the output, and they generally show higher accuracy than decision trees, also being apt to handle missing and incomplete data, without many shortcomings about the data type.

In our approach, we resorted to REVAC (Montero et al. 2014), which is an Estimation of Distribution Algorithm that implements some concepts stemming from information theory to measure parameter relevance, via considering the parameters' value distributions over the parameter space, and assigns high probabilities to values that lead to a good compromise between the algorithm's performance and Shannon entropy (i.e., its complexity): the algorithm creates a uniform distribution over the parameters space, which is iteratively smoothed via an evolutionary process whose individuals correspond to parameter vectors, and whose fitness is evaluated by their expected performance.

4.3 Artificial neural networks

Artificial Neural Networks (ANN) are high-level computational tools inspired to the behavior of the brain, of which they represent a simplification: they are composed of elementary units (i.e., neuron, to whom a numerical value is associated) and weighted oriented hedges that connect them (i.e., the synapses, whose value is modified over time by the learning algorithm): these weights are subject to learning that may occur in different ways (reinforcement, supervised, and non-supervised). Decision trees, Random forests, and Neural Networks need a training phase, in which the learning phase is performed: this is generally performed on a specific set

of data, referred to as training set, while the algorithm is normally tested (or the learning is evaluated) on a different set of data, referred to as test set. In our experiments, we employed feed-forward networks with Backpropagation momentum, utilizing REVAC to determine the learning and momentum rates. For establishing the network topology, we adopted the adaptive procedure developed by Corazza et al. (2021). This procedure starts with a network consisting of one neuron in the hidden layer and progressively adds neurons to the hidden layer until no improvement is observed over a user-defined number of iterations on the test set, aiming to minimize the calculated error.

5 Our computational approach

The objective of this article is to propose a computational approach for predicting prices based on listings-based data. In pursuit of this goal, we compare various tools, including linear models, generalized linear models, decision trees, random forests, and neural networks. From an operational point of view, experiments have been run on a laptop equipped with 16 GB RAM and an Intel Core Intel 2.80 GHz CPU, and in what follows, Sect. 5.1 presents the available data, which will undergo significant pre-processing operations useful for subsequent analyses. In Sect. 5.2.1, we introduce the computational tools covering linear regression and generalized linear models. In Sect. 5.2.2, we delve into decision trees. random forests and neural networks.

5.1 Our data

We have used data about all Airbnb listings in the Netherlands operating on Summer 2019. Data have been provided to us by a commercial service that uses web scraping techniques to collect Airbnb usage information from the Airbnb website daily (named AirDNA), and consists of features of 142,851 listings. The scraping procedure explores two levels in the page tree whose root is given by the single listing's page, and the resulting set of data includes all active Airbnb and HomeAway listings in the Netherlands as of September 2019, presenting comprehensive details such as geographic locations, performance statistics, including average prices, number of bookings and review counts. Additionally, the dataset provides information on Airbnb's monthly activity, including start date, up to 2013. We dispose of data about all features reported in Table 2, that also reports the main statistics about data that will be used as predictors in our approach. Numbers have been truncated after two decimal digits. In this study, data were not winsorized, since (i) no especially anomalous values were found in a preliminary outlier detection analysis based on z-scores, and (ii) no particular valid justification was found for removing observations. Once the data is grouped by city and lines with missing entries are removed, extreme values exceed three standard deviations away from the mean -a standard practice in outlier detection analysis for detecting 99% of normally distributed data- in less than 10% of the cities included in the sample. As suggested by Sainaghi & Baggio (2020), the Average

Table 2 Main statistics of Airbnb listings' features from the whole set of data, used as predictors in the models proposed in what follows

Variable name	Mean	SD	Min	Max
Annual revenue LTM (USD)	7060.45	15,346.97	0	507,683
Occupancy rate LTM	0.33	0.37907	0	22
Number of bookings LTM	11.42	21.73	0	230
Number of reviews	12.97	30.41	0	733
Bedrooms	1.63	1.22	0	40
Bathrooms	1.139159	0.66	0	40
Max guests	3.44	2.38	0	98
Response rate	72.46	41.11	0	100
Security deposit (USD)	110.12	294.90	0	11,338
Cleaning fee (USD)	26.28	35.97	0	1274
Extra people Fee (USD)	8.38	19.23	0	346
Minimum stay	3.26	16.71	0	3999
Count reservation days LTM	47.11	73.41	0	365
Count available days LTM	26.70	46.80	0	325
Count blocked days LTM	16.21	27.63	0	254
Number of photos	14.77	10.73	0	201
Overall rating	5.53	4.57	0	10
Airbnb Communication rating	5.79	4.78	0	10
Airbnb accuracy rating	5.68	4.72	0	10
Airbnb cleanliness rating	5.56	4.62	0	10
Airbnb checking rating	5.78	4.78	0	10
Airbnb location rating	5.46	4.54	0	10

Daily Rate will be used as the dependent variable of our predicting exercise, and we have used it as measured in USD. For our exercise, we have only used predictors directly coming from the Airbnb listing, without any knowledge coming from the environment, e.g., the number of hotels and/or supermarkets in the neighborhood, the distance from city-center, the population density of the neighborhood, just to name the most used. As we have already pointed out in Sect. 2, we are introducing some variables that represent the listing's (and host's) historic performances: the last twelve months (referred to as LTM) revenue, the occupancy rate, the number of bookings, the count of reservations, of the available days on the agenda, and the number of blocked days: this is important because most literature does not take into account the financial history of the listing. Furthermore, we have not used any categorical variables to avoid conversion problems, nor made any assumption about competition (Chen and Xie 2017) and related measures and thresholds (Cai et al. 2019a).

Observing statistical features of the different geographical locations leads to interesting preliminary considerations: not surprisingly, the city of Amsterdam hosts the listings with the highest occupancy rate and annual revenue, but also the

ones with the lowest occupancy rate and annual revenue, which is a sign of the great diversification of the Airbnb offer in that city (59,099 listings). The capital city also features the listing with maximum (and minimum) values of cleaning fee, number of review extra people fee, and number of photos. Interestingly instead, the highest average daily rate is found in the city of Rotterdam, and the minimum in Den Haag: this latter city also features the listing with the highest number of bookings over the year. It is also worthwhile to mention the city of Eindhoven, which features the listings with the maximum allowed number of guests, minimum reported stay, and available days over year. High number of reservations are also found in the city of Castricum (seaside town in North Holland, attractive for the beach and nearby dune landscape), and high reservation days (over one year) are in the city of Bergen, whose beaches and forests make it a popular destination for tourists; high occupancy rate is shown by the city of Arnhem, probably due to its closeness to the German frontier. Anyhow, we can state that the presence of Airbnb in the Netherlands is quite widespread and found its application also in smaller and less touristic attractive locations.

5.2 Computational results

In this section, we present the main results of our experiments, along with a discussion of the relationship between the distinctive characteristics of the models and the very nature of the problem at hand. In the following, we take care in particular of model assumptions and the added value brought by the inclusion of the financial history into the predictive model.

5.2.1 Linear regression and generalized linear models

As a first experiment, we have implemented a standard linear regression and a GLM approach, by using the Average Daily Rate as dependent variable, and all variables outlined in Table 2 as regressors. Please notice that, to avoid collinearity related issues, we have decided to not use variable Overall Rating, that is often a combination of variables Airbnb Communication Rating, Airbnb Accuracy Rating, Airbnb Cleanliness Rating, Airbnb Check-in Rating, and Airbnb Location Rating.

Preliminary experiments have shown a good significance of the linear model, seeming to suggest that a linear trend is present on the data, since the regression was highly statistically significant (p-value of the F -test lower than 0.001, and a R^2 equal to 0.642, see what follows), but we have performed an analysis of the residual of the regression, leading to remark that the residuals standard deviations is proportional to the fitted values for all cities taken into account, since the ratio residuals standard deviations to mean of the fitted values is roughly constant, suggesting heteroscedasticity in the data, and hence, to adopt a logarithmic transformation to reduce heteroscedasticity and to provide more robust estimates. In view of these observations, we have used a logarithm transformation of all variables considered. Table 3 reports the estimates of the parameters of each variable taken into account, along with their p-values (in a nutshell, a p-value smaller than 0.05 indicates that the

Table 3 Predictor coefficients and p-values (values lower than 0.05 indicate that the corresponding predictor is statistically significant) obtained with linear regression and Generalized Linear Model with Gamma variance function and inverse link function, along with the main regression's performance measures, obtained on the whole data

Variable name	Linear model		Generalized linear model	
	Coefficient	*p*-values	Coefficient	*p*-values
Intercept	− 1.047	**< 0.01**	− 1.270	**< 0.01**
Annual Revenue LTM (USD)	0.006	**< 0.01**	**< 0.001**	**< 0.01**
Occupancy Rate LTM	163.221	**< 0.01**	**< 0.001**	0.073
Number of Bookings LTM	0.311	**< 0.01**	0.014	**< 0.01**
Number of Reviews	0.071	**< 0.01**	0.005	**< 0.01**
Bedrooms	5.384	**< 0.01**	− 0.016	0.09
Bathrooms	7.361	**< 0.01**	− 0.127	**< 0.01**
Max Guests	0.497	0.001	0.002	0.70
Response Rate	− 0.013	0.023	0.003	**< 0.01**
Security Deposit (USD)	0.009	**< 0.01**	**< 0.001**	**< 0.01**
Cleaning Fee (USD)	0.241	**< 0.01**	− 0.004	**< 0.01**
Extra People Fee (USD)	0.084	**< 0.01**	− 0.001	**< 0.01**
Minimum Stay	− 0.0020	0.871	**< 0.001**	0.87
Count Reservation Days LTM	− 1.191	**< 0.01**	0.026	**< 0.01**
Count Available Days LTM	0.692	**< 0.01**	− 0.011	**< 0.01**
Count Blocked Days LTM	0.150	**< 0.01**	0.019	**< 0.01**
Number of Photos	0.138	**< 0.01**	**< 0.001**	**< 0.01**
Airbnb Communication Rating	0.1559	0.696	− 0.002	0.866
Airbnb Accuracy Rating	1.4241	**< 0.01**	0.032	**0.017**
Airbnb Cleanliness Rating	− 0.0607	0.889	− 0.023	**0.026**
Airbnb Checkin Rating	2.3285	**< 0.01**	− 0.009	0.501
Airbnb Location Rating	− 3.0052	**< 0.01**	− 0.023	**0.029**
Statistics				
R^2	0.684		0.723	
Adjusted R2	0.684		0.642	
p-value of F-test	< 0.01		< 0.01	

variable is significant in the regression model), for both the Linear and Generalised Linear models. In addition, in order to assess the significance of the model, we have reported the coefficient of determination R^2, along with its Adjusted version, that takes into account the number of variables and observation, and we can see that there is no difference between these two values. We have also reported the *p*-value of the F -test, that represents the probability to obtain a F-statistic value greater than the F-value of the model, under the null hypothesis that the regression is not significant (in a nutshell, values smaller than 0.05 lead to reject this null hypothesis, hence confirming the significance of the model). For the sake of readability, *p*-values smaller than 0.05 are highlighted in bold.

We can see that almost all variables are significant in the regression, with some notable exceptions: the Cleanliness rating does not have significance on both models, suggesting that guests do not recognize the user-experience as criterion to an overnight stay on Airbnb. Also, the minimum stay indicated by the host has no effect on pricing, and this could be given by the low entity of cleaning fee, that could have a limited effect on the user experience: this aspect has to be investigated in further works. We remark anyhow that the number of variables significant in the linear model (19) is greater than the one related to the GLM (16), indicating that GLM may perform well using a fewer number of predictors: this goes in the same direction of the analysis of R^2, which his higher when using GLM: furthermore, both models are significant according to the F-test, hence both methods reveal that it is possible to model the pricing exercise and to obtain satisfactory results by using standard tools as linear and generalized linear models: this seems to suggest that some regularities arise in the pricing mechanism, that, at least on the Netherlands, seems to be apt to be modelled when using both listings-related and historical-performances data.

5.2.2 Evaluation of predictions

In order to test the robustness of our approaches, following the approach by Angelini et al. (2008), we partitioned the data at hand into train and test (hold-out) according to the 70/30 percent rule. This process has been repeated 30 times, in order to test our approaches over 30 different train-test partitions, and for each partition, the performance metrics have been assessed over the test set. Table 4 reports the main statistics of the R2 and the Mean Squared Error of each method over the 30 different test sets.

It is noteworthy that Random Forests exhibit superior performance, demonstrating low variability in results, as indicated by low standard deviations. They consistently outperform individual decision trees. Interestingly, standard tools like Generalized Linear Models (GLMs) offer performances on par with more complex black-box models such as neural networks. Linear models, although yielding

Table 4 Main performance statistics for different methods of prices prediction employed on the instance containing all listings from the overall country

Variable name	R^2				MSE		
	Average	STD	Min	Max	Average	STD	Min
Random forest	0.76	0.04	0.71	0.79	8.85	2.89	6.96
Decision tree	0.63	0.52	0.21	0.91	9.34	3.05	7.41
Neural networks	0.71	0.15	0.62	0.88	3.12	5.01	2.91
Linear regression	0.64	0.23	0.51	0.72	9.12	3.25	8.43
GLM	0.71	0.19	0.64	0.78	9.01	2.77	8.61

All measures are relative to the test sets

Table 5 Main statistics for different methods of price prediction across different cities and the whole country

Listings from	Random forests	Neural networks	Linear regression	GLM	Ensemble
Alphen aan Den Rijn	0.99	0.80	0.64	0.65	0.71
Amsterdam	0.78	0.76	0.72	0.75	0.89
Bussum	0.79	0.75	0.58	0.72	0.48
De Wolden	0.79	0.73	0.26	0.61	0.56
Delft	0.74	0.79	0.41	0.54	0.53
Eindhoven	0.62	0.71	0.39	0.59	0.59
Leiden	0.59	0.93	0.81	0.44	0.59
Rotterdam	0.72	0.45	0.55	0.60	0.63
Texel	0.65	0.70	0.63	0.69	0.79
The Hague	0.75	0.68	0.36	0.51	0.76
Tilburg	0.82	0.92	0.21	0.85	0.78
Vlieland	0.92	0.91	0.35	0.86	0.85
Zwolle	0.49	0.81	0.69	0.52	0.79
Netherlands	0.76	0.71	0.64	0.71	0.68

All measures are relative to the test sets

satisfactory results based on the introduced measures, exhibit the weakest performance among the analyzed techniques.

An intriguing avenue for exploration involves comparing the results obtained for listings from the entire Netherlands with those from individual cities. Table 5 presents the R^2 values obtained by the aforementioned techniques on a selected sample of cities in the Netherlands. Please note that, despite providing us with a handy evaluation of models, the assessment of the R2 score on the validation set should not be interpreted as an evaluation of the goodness-of-fit of models across cities. In general, the same observations from the previous paragraph hold: Random Forests demonstrate superior results, GLM performances are comparable to Neural Networks, and linear models, while satisfactory, lag behind the other techniques. However, some variations emerge in specific cities (Table 5). For instance, Neural Networks outperform Random Forests in various cities, including Leiden, Delft, and Eindhoven. Additionally, linear models exhibit stronger performances than GLMs in cities such as Leiden and Zwolle. Notably, Random Forests do not deliver satisfactory results in the latter city.

A voting ensemble approach combining the prediction from multiple models (linear regression, random forests and neural networks) is then assessed. It simply amounts to averaging the individual predictions of different estimators to obtain a final prediction, in order to check whether improvements in the performance of individual learners can be attained, using a straightforward combination of their predictions. The results it yields are good, yet not clearly outperforming. This could be related to the negative impact of the linear regression model on the overall performance of the ensemble learner, since it systematically underperforms other alternatives. Consequently, we draw the following conclusions: the linear may be valuable

for assessing the significance of different predictors, but for predictive tasks, a more effective strategy involves the combined use of diverse intelligent techniques. Additionally, our approach demonstrates efficiency, with the overall computational time below five minutes on a common computer architecture.

Given the diverse nature and tourist orientation of the cities under consideration, an immediate comparison between them is not straightforward. However, it is worth noting that, on the one hand, cities of significant importance, albeit not exclusively specialized in tourism, such as Amsterdam, demonstrate commendable performance across all the models employed. On the other hand, medium-sized and less open-to-tourism cities, such as Delft, Eindhoven and The Hague, show superior performance in random forest and neural network models compared to their more traditional counterparts. Conducting further analyses on the potential implications of city typologies, their specialization in tourism, and the nature of tourist accommodation—whether traditional with hotels or non-traditional with rentals, even by non-professionals—could therefore be instrumental in understanding whether specific characteristics of cities are associated with optimal performance in a particular forecasting model.

Finally, we reconsider the role of newly introduced features from a machine learning point of view, by assessing each variable's contribution based on Shapley values (Lundberg and Lee 2017) and permutation importance (Breiman 2001), respectively, as shown in Figs. 1 and 2. Although both methods are model-agnostic, for illustrative purposes, we limit ourselves to gauging the effect w.r.t random forests. First, observe that ratings and reviews have a limited impact on the prediction
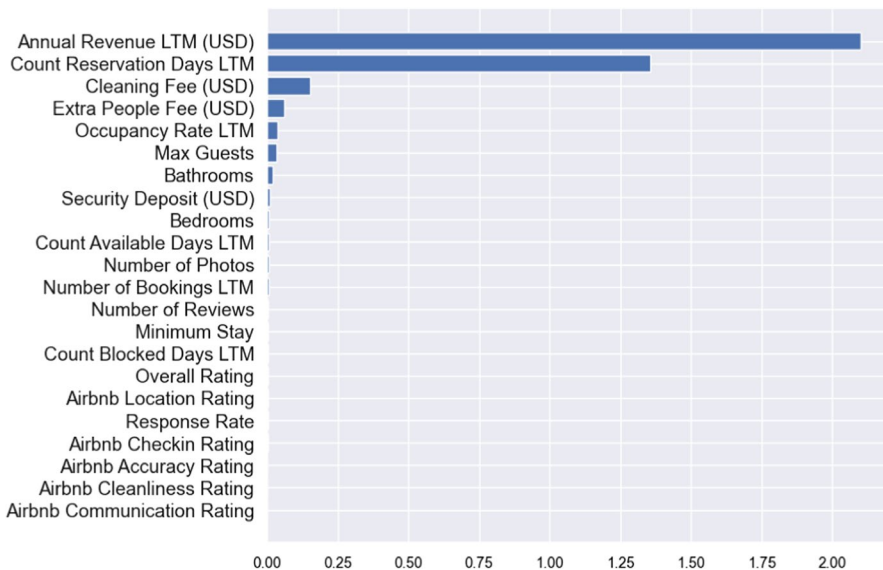
Fig. 1 Permutation feature importance based on random forests evaluations on test data. The size of the bars represents the importance of a feature, i.e. the increase in the prediction error of the model after the feature's values are permuted
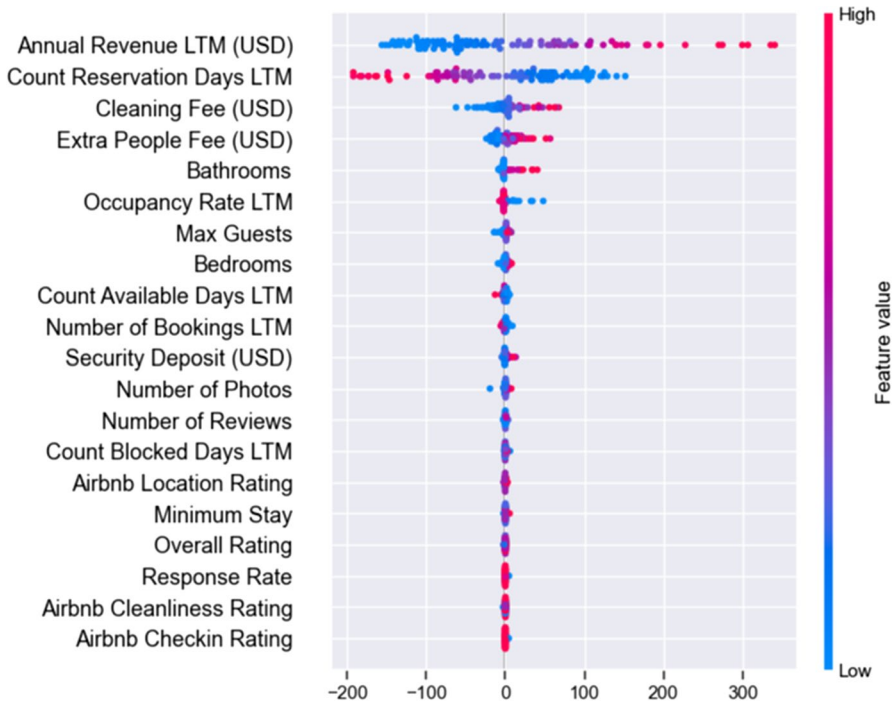
Fig. 2 Shapley values (impact on model output) based on random forests evaluations on test data. The x-axis reports the Shapley values, while the colors denote the size of the features, according to the scale reported in the colorbar on the right of the plot (color figure online)

performance of the model, whereas the financial features of the listings are by far the most relevant. The consequences of these observations are, at least, threefold: (i) we corroborate the evidence that the past users' feedback, despite being highly visible to potential customers, has limited impact; (ii) similarly, the ratings provided on the website, regardless of accessibility to potential users, attain little importance; (iii) on the contrary, the last twelve months performance, along with other relevant fees, are critical for predictive purposes.

6 Conclusion

In this article, we have devised various methods for predicting Airbnb prices, encompassing both conventional approaches (linear and generalized linear models) and Artificial Intelligence techniques (random forest, decision trees, and neural networks). Our analysis is grounded in real data sourced from the Netherlands, specifically focusing on all Airbnb listings active until September 2019. Distinguishing our study from others, we have incorporated the financial history of diverse rental offerings as part of the analyzed characteristics. Additionally, our

investigation extends beyond individual cities or regions, encompassing a broader spectrum of cities across the Netherlands.

Results show that Artificial Intelligence based methods have more robust predicting skills and are less prone to over-fitting, and they outperform more traditional approaches over the introduced performance metrics. On the other hand, results obtained by traditional methods are easier to interpret, also in view of determining the significance of the predictors in the predicting exercise. However, we have remarked good predicting skills of all methods, and this is important with respect to the specific research question we have devised in the introduction: the non-professional pricing mechanism applied by many Airbnb hosts led to some doubts about the efficiency (and suitability) of using predicting price tools, due to the fact that Airbnb price could be determined in a non-professional way: in our contribution, instead, we have seen that linear models show good predicting skills, determining a good deterministic mathematical relation between price and a set of predictors, but also that Artificial Intelligence based methods are not sensitive to this consideration, and that they are able to perform a successful pricing exercise over the data at hand. In both cases, variables that represent the listing's (and host's) historic performances, that have been somewhat neglected so far, have a significant prediction impact, and their use, along with the listing features, location-based attributes, social- and platform-based- indicators can be successfully used to predict Airbnb prices. Further analyses of Airbnb pricing should deal with some of the limitations of our work. For instance, in this contribution, a specific array of models has been considered; a special focus on Deep Learning would allow a better understanding of the potential of Artificial Intelligence in this context and possibly improve prediction accuracy. Further efforts should focus on boosting algorithms that are known to perform particularly well on tabular data, such as LightGBM or XGBoost. Moreover, our analysis is limited to the Netherlands; hence, our conclusions rely on the assumption that our sample constitutes a good representation of the universe of Airbnb listings. At this stage, cross-country heterogeneity is a feature that we do not control for here, and that should be taken into account for future developments. Possible future works might include an adaptive selection of the computational tool to be used during the prediction exercise, on the basis of performance metrics found during previous applications, as suggested by Filograsso and di Tollo (2023): of course, this will necessitate the definition of some metrics to be used to this extent.

Furthermore, a joint use of the techniques investigated in this contribution should be encouraged, through the definition of suitable aggregation operators, to provide a unified framework, able to model price and give some recommendations to non-professional hosts.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Adamiak C (2022) Current state and development of Airbnb accommodation offer in 167 countries. Curr Issue Tour 25(19):3131–3149. https://doi.org/10.1080/13683500.2019.1696758

Adamiak C, Szyda B, Dubownik A, García-Álvarez D (2019) Airbnb offer in Spain—spatial analysis of the pattern and determinants of its distribution. ISPRS Int J Geo Inf 8(3):155. https://doi.org/10.3390/ijgi8030155

Adamiak, C. (2018). Mapping Airbnb supply in European cities.

AirDNA, LLC. (2018). Short-term rental data methodology - the AI behind AirDNA. Retrieved from AirDNA: https://www.airdna.co/blog/short-term-rental-datamethodology

Albors J, Ramos JC, Hervas JL (2008) New learning network paradigms: communities of objectives, crowdsourcing, wikis and open source. Int J Inf Manage 28(3):194–202. https://doi.org/10.1016/j.ijinfomgt.2007.09.006

Andria J, di Tollo G, Pesenti R (2021) Fuzzy multi-criteria decision-making: an entropy-based approach to assess tourism sustainability. Tour Econ 27(1):168–186. https://doi.org/10.1177/1354816619885207

Angelini E, Di Tollo G, Roli A (2008) A neural network approach for credit risk evaluation. Q Rev Econ Finance 48(4):733–755. https://doi.org/10.1016/j.qref.2007.04.001

Anselin L (1988) Spatial econometrics: methods and models. Springer Science & Business Media, Dordrecht

Ball, J., Arnett, G., & Franklin, W. (2014). London's Buy-to-let Landlords Look to Move in on Spare Room Website Airbnb'. *The Guardian*, 20.

Bernardi, M. (2018). The impact of AirBnB on our cities: Gentrification and "disneyfication" 2.0. LabGov. *City, The Urban Media Lab*.

Van der Borg, J., Camatti, N., Bertocchi, D., & Albarea, A. (2017). The rise of the sharing economy in tourism: Exploring Airbnb attributes for the Veneto Region. *University Ca'Foscari of Venice, Dept. of Economics Research Paper Series No*, 5. https://doi.org/10.2139/ssrn.2997985

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Cai Y, Zhou Y, Scott N (2019a) Price determinants of Airbnb listings: evidence from Hong Kong. Tour Anal 24(2):227–242. https://doi.org/10.3727/108354219x15525055915554

Cai T, Han K, Wu H (2019b) Melbourne airbnb price prediction

Čeh M, Kilibarda M, Lisec A, Bajat B (2018) Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. ISPRS Int J Geo Inf 7(5):168. https://doi.org/10.3390/ijgi7050168

Chattopadhyay M, Mitra SK (2019) Do airbnb host listing attributes influence room pricing homogenously? Int J Hosp Manag 81:54–64. https://doi.org/10.1016/j.ijhm.2019.03.008

Chau KW, Chin TL (2003) A critical review of literature on the hedonic price model. Int J Hous Sci Appl 27(2):145–165

Chen Y, Xie K (2017) Consumer valuation of Airbnb listings: A hedonic pricing approach. Int J Contemp Hosp Manag 29(9):2405–2424. https://doi.org/10.1108/ijchm-10-2016-0606

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: online book reviews. J Mark Res 43(3):345–354. https://doi.org/10.1509/jmkr.43.3.345

Chica-Olmo J, González-Morales JG, Zafra-Gómez JL (2020) Effects of location on Airbnb apartment pricing in Málaga. Tour Manage 77:103981. https://doi.org/10.1016/j.tourman.2019.103981

Corazza M, De March D, di Tollo G (2021) Design of adaptive Elman networks for credit risk assessment. Quant Financ 21(2):323–340. https://doi.org/10.1080/14697688.2020.1778175

Dana JD Jr (2008) New directions in revenue management research. Prod Oper Manag 17(4):399–401. https://doi.org/10.3401/poms.1080.0040

Deboosere R, Kerrigan DJ, Wachsmuth D, El-Geneidy A (2019) Location, location and professionalization: a multilevel hedonic analysis of Airbnb listing prices and revenue. Reg Stud Reg Sci 6(1):143–156. https://doi.org/10.1080/21681376.2019.1592699

Dolnicar S (2019) A review of research into paid online peer-to-peer accommodation: launching the annals of tourism research curated collection on peer-to-peer accommodation. Ann Tour Res 75:248–264. https://doi.org/10.1016/j.annals.2019.02.003

Dudás G, Vida G, Kovalcsik T, Boros L (2017) A socio-economic analysis of Airbnb in New York City. Reg Stat 7(1):135–151. https://doi.org/10.15196/rs07108

Ert E, Fleischer A, Magen N (2016) Trust and reputation in the sharing economy: the role of personal photos in Airbnb. Tour Manage 55:62–73. https://doi.org/10.1016/j.tourman.2016.01.013

Eugenio-Martin JL, Cazorla-Artiles JM, González-Martel C (2019) On the determinants of Airbnb location and its spatial distribution. Tour Econ 25(8):1224–1244. https://doi.org/10.1177/1354816618825415

Fagerstrøm A, Pawar S, Sigurdsson V, Foxall GR, Yani-de-Soriano M (2017) That personal profile image might jeopardize your rental opportunity! On the relative impact of the seller's facial expressions upon buying behavior on Airbnb™. Comput Hum Behav 72:123–131. https://doi.org/10.1016/j.chb.2017.02.029

Falk M, Larpin B, Scaglione M (2019) The role of specific attributes in determining prices of Airbnb listings in rural and urban locations. Int J Hosp Manag 83:132–140. https://doi.org/10.1016/j.ijhm.2019.04.023

Filograsso G, di Tollo G (2023) Adaptive evolutionary algorithms for portfolio selection problems. Comput Manag Sci. https://doi.org/10.1007/s10287-023-00441-7

Fu S, Cheng X, Bao Y, Bilgihan A, Okumus F (2021) Staying in a hotel or peer-to-peer accommodation sharing? a discrete choice experiment with online reviews and discount strategies. Internet Res 31(2):654–676. https://doi.org/10.1108/intr-01-2020-0031

Gibbs C, Guttentag D, Gretzel U, Morton J, Goodwill A (2018a) Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings. J Travel Tour Mark 35(1):46–56. https://doi.org/10.1080/10548408.2017.1308292

Gibbs C, Guttentag D, Gretzel U, Yao L, Morton J (2018b) Use of dynamic pricing strategies by Airbnb hosts. Int J Contemp Hosp Manag 30(1):2–20. https://doi.org/10.1108/ijchm-09-2016-0540

Grant, M. (2013). Airbnb. com poses only a small threat to hotel industry. Euromonitor International.

Gunter U, Önder I (2018) Determinants of Airbnb demand in Vienna and their implications for the traditional accommodation industry. Tour Econ 24(3):270–293. https://doi.org/10.1177/1354816617731196

Gutt, D., & Herrmann, P. (2015, May). Sharing Means Caring? Hosts' Price Reaction to Rating Visibility. In *ECIS* (Vol. 54).

Guttentag D (2015) Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. Curr Issue Tour 18(12):1192–1217. https://doi.org/10.1080/13683500.2013.827159

Guttentag D (2019) Progress on Airbnb: a literature review. J Hosp Tour Technol 10(4):814–844. https://doi.org/10.1108/jhtt-08-2018-0075

Guttentag D, Smith S, Potwarka L, Havitz M (2018) Why tourists choose Airbnb: a motivation-based segmentation study. J Travel Res 57(3):342–359. https://doi.org/10.1177/0047287517696980

Hamari J, Sjöklint M, Ukkonen A (2016) The sharing economy: Why people participate in collaborative consumption. J Am Soc Inf Sci 67(9):2047–2059. https://doi.org/10.1002/asi.23552

Harewood SI (2006) Managing a hotel's perishable inventory using bid prices. Int J Oper Prod Manag 26(10):1108–1122. https://doi.org/10.1108/01443570610691094

Hill D (2015) How much is your spare room worth? IEEE Spectr 52(9):32–58

Hossain M (2020) Sharing economy: a comprehensive literature review. Int J Hosp Manag 87:102470. https://doi.org/10.1016/j.ijhm.2020.102470

Hung WT, Shang JK, Wang FC (2010) Pricing determinants in the hotel industry: quantile regression analysis. Int J Hosp Manag 29(3):378–384. https://doi.org/10.1016/j.ijhm.2009.09.001

Ikkala, T., & Lampinen, A. (2014). Defining the price of hospitality. In: Proceedings of the companion publication of the 17th ACM conference on computer supported cooperative work & social computing. https://doi.org/10.1145/2556420.2556506

Johnson MP (2003) Single-period location models for subsidized housing: tenant-based subsidies. Ann Oper Res 123:105–124. https://doi.org/10.1016/j.seps.2004.11.001

Karlsson L, Dolnicar S (2016) Someone's been sleeping in my bed. Ann Tour Res 58:159–162

Kokasih, M. F., & Paramita, A. S. (2020). Property rental price prediction using the extreme gradient boosting algorithm. https://doi.org/10.47738/ijiis.v3i2.65

Kwok L, Xie KL (2019) Pricing strategies on Airbnb: Are multi-unit hosts revenue pros? Int J Hosp Manag 82:252–259. https://doi.org/10.1016/j.ijhm.2018.09.013

Lampinen, A., & Cheshire, C. (2016, May). Hosting via Airbnb: Motivations and financial assurances in monetized network hospitality. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 1669–1680). https://doi.org/10.1145/2858036.2858092

Lawani A, Reed MR, Mark T, Zheng Y (2019) Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston. Reg Sci Urban Econ 75:22–34. https://doi.org/10.1016/j.regsciurbeco.2018.11.003

Li, Y., Pan, Q., Yang, T., & Guo, L. (2016). Reasonable price recommendation on Airbnb using Multi-Scale clustering. In: *2016 35th Chinese control conference (CCC)* (pp. 7038–7041). IEEE. https://doi.org/10.1109/chicc.2016.7554467

Limsombunchao, V. (2004). House price prediction: hedonic price model vs. artificial neural network. https://doi.org/10.3844/ajassp.2004.193.201

Lorde T, Jacob J, Weekes Q (2019) Price-setting behavior in a tourism sharing economy accommodation market: a hedonic price analysis of AirBnB hosts in the caribbean. Tour Manag Perspect 30:251–261. https://doi.org/10.1016/j.tmp.2019.03.006

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Luo Y, Zhou X, Zhou Y (2019) Predicting Airbnb listing price across different cities. Highlights Sci Eng Technol. https://doi.org/10.54097/hset.v47i.8169

McAfee RP, Te Velde V (2006) Dynamic pricing in the airline industry. Handbook on economics and information systems 1:527–67

Mermet, A. C. (2019). The 'Airbnbfication' of the Icelandic capital. Towards an assessment of the socio-spatial impacts of Airbnb in the Reykjavik capital area (Doctoral dissertation, Ferðamálastofa).

Mikhalkina, T., & Cabantous, L. (2015). Business model innovation: How iconic business models emerge. In *Business models and modelling* (pp. 59–95). Emerald Group Publishing Limited. https://doi.org/10.1108/s0742-332220150000033024

Montero E, Riff M-C, Neveu B (2014) A beginner's guide to tuning methods. Appl Soft Comput 17:39–51

Moreno-Izquierdo L, Ramón-Rodríguez AB, Such-Devesa MJ, Perles-Ribes JF (2019) Tourist environment and online reputation as a generator of added value in the sharing economy: the case of Airbnb in urban and sun- and-beach holiday destinations. J Destin Mark Manag 11:53–66. https://doi.org/10.1016/j.jdmm.2018.11.004

Önder I, Weismayer C, Gunter U (2018) Spatial price dependencies between the traditional accommodation sector and the sharing economy. Tour Econ 25(8):1150–1166. https://doi.org/10.1177/1354816618805860

Oskam, J. A. (2019). The future of Airbnb and the 'sharing economy': The collaborative consumption of our cities (Vol. 1). Channel View Publications. https://doi.org/10.21832/9781845416744

Park B, Bae JK (2015) Using machine learning algorithms for housing price prediction: the case of Fairfax County, Virginia housing data. Expert Syst Appl 42(6):2928–2934. https://doi.org/10.1016/j.eswa.2014.11.040

Perez-Sanchez VR, Serrano-Estrada L, Marti P, Mora-Garcia RT (2018) The what, where, and why of Airbnb price determinants. Sustainability 10(12):4596. https://doi.org/10.3390/su10124596

Quattrone G, Greatorex A, Quercia D, Capra L, Musolesi M (2018) Analyzing and predicting the spatial penetration of Airbnb in US cities. EPJ Data Sci 7(1):31. https://doi.org/10.1140/epjds/s13688-018-0156-6

Rezazadeh Kalehbasti, P., Nikolenko, L., & Rezaei, H. (2021, August). Airbnb price prediction using machine learning and sentiment analysis. In: *international cross-domain conference for machine learning and knowledge extraction* (pp. 173–184). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-84060-0_11

Rubino, I., & Coscia, C. (2018, May). Airbnb revenue generation in the urban context: An analysis of renting patterns and dynamics. In *international symposium on new metropolitan perspectives* (pp. 643–651). Cham: Springer International Publishing.

Sainaghi R (2020) The current state of academic research into peer-to-peer accommodation platforms. Int J Hosp Manag 89:102555. https://doi.org/10.1016/j.ijhm.2020.102555

Sainaghi R (2021) Determinants of price and revenue for peer-to-peer hosts The state of the art. Int J Contemporary Hospitality Manag 33(2):557–586. https://doi.org/10.1108/ijchm-08-2020-0884

Sainaghi R, Baggio R (2020) Clusters of topics and research designs in peer-to-peer accommodation platforms. Int J Hosp Manag 88:102393. https://doi.org/10.1016/j.ijhm.2019.102393

Schelling TC (1969) Models of segregation. Am Econ Rev 59(2):488–493

Schor JB, Attwood-Charles W (2017) The "sharing" economy: labor, inequality, and social connection on for-profit platforms. Sociol Compass 11(8):e12493. https://doi.org/10.1111/soc4.12493

Smith GS, Thorpe RJ Jr (2020) Gentrification: a priority for environmental justice and health equity research. Ethnicity Dis 30(3):509. https://doi.org/10.18865/ed.30.3.509

Stabrowski F (2017) 'People as businesses': Airbnb and urban micro-entrepreneurialism in New York City. Camb J Reg Econ Soc 10(2):327–347. https://doi.org/10.1093/cjres/rsx004

Sthapit E, Jiménez-Barreto J (2019) You never know what you will get in an Airbnb: poor communication destroys value for guests. Curr Issue Tour 22(19):2315–2318. https://doi.org/10.1080/13683500.2018.1475469

Sundararajan A (2017) The sharing economy: The end of employment and the rise of crowd-based capitalism. MIT Press. https://doi.org/10.3917/rimhe.035.0092

Sutherland W, Jarrahi MH (2018) The sharing economy and digital platforms: a review and research agenda. Int J Inf Manage 43:328–341. https://doi.org/10.1016/j.ijinfomgt.2018.07.004

Tang, E., & Sangani, K. (2015). Neighborhood and price prediction for San Francisco Airbnb listings. *Departments of Computer science, Psychology, economics–Stanford University*. https://doi.org/10.24148/cdwp2021-01

Tang LR, Kim J, Wang X (2019) Estimating spatial effects on peer-to-peer accommodation prices: toward an innovative hedonic model approach. Int J Hospital Manag 81:43–53. https://doi.org/10.1016/j.ijhm.2019.03.012

Teubner T, Hawlitschek F, Dann D (2017) Price determinants on Airbnb: How reputation pays off in the sharing economy. J Self-Govern Manag Econ 5(4):53–80. https://doi.org/10.22381/jsme5420173

Tong B, Gunter U (2020) Hedonic pricing and the sharing economy: how profile characteristics affect Airbnb accommodation prices in Barcelona, Madrid, and Seville. Curr Issue Tour 25(20):3309–3328. https://doi.org/10.1080/13683500.2020.1718619

Türk U, Östh J, Kourtit K, Nijkamp P (2021) The path of least resistance explaining tourist mobility patterns in destination areas using Airbnb data. J Transp Geogr 94:103130. https://doi.org/10.1016/j.jtrangeo.2021.103130

Visser G, Erasmus I, Miller M (2017) Airbnb: the emergence of a new accommodation type in Cape Town. South Africa Tour Rev Int 21(2):151–168. https://doi.org/10.3727/154427217X14912408849458

Wang D, Nicolau JL (2017) Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb. com. Int J Hosp Manag 62:120–131. https://doi.org/10.1016/j.ijhm.2016.12.007

Wegmann J, Jiao J (2017) Taming Airbnb: Toward guiding principles for local regulation of urban vacation rentals based on empirical results from five US cities. Land Use Policy 69:494–501. https://doi.org/10.1016/j.landusepol.2017.09.025

Xie KL, Kwok L, Heo CY (2019a) Are neighbors friends or foes? assessing Airbnb listings' agglomeration effect in New York City. Cornell Hospital Q 61(2):128–141. https://doi.org/10.1177/1938965519890578

Xie K, Mao Z, & Wu, J. (2019b) Learning from peers: The effect of sales history disclosure on peer-to-peer short-term rental purchases. Int J Hosp Manag 76:173–183. https://doi.org/10.1016/j.ijhm.2018.05.007

Xu F, Hu M, La L, Wang J, Huang C (2019) The influence of neighbourhood environment on Airbnb: a geographically weighed regression analysis. Tour Geogr. https://doi.org/10.1080/14616688.2019.1586987

Yao B, Qiu RT, Fan DX, Liu A, Buhalis D (2019) Standing out from the crowd–an exploration of signal attributes of Airbnb listings. Int J Contemp Hosp Manag 31(12):4520–4542. https://doi.org/10.1108/ijchm-02-2019-0106

Zervas G, Proserpio D, Byers JW (2017) The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. J Mark Res 54(5):687–705. https://doi.org/10.1509/jmr.15.0204

Zhang Z, Chen RJ, Han LD, Yang L (2017) Key factors affecting the price of Airbnb listings: A geographically weighted approach. Sustainability 9(9):1635. https://doi.org/10.3390/su9091635

Publisher's Note   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.